Carnegie Foundation
for the Advancement of Teaching

JANUARY
2014

ISSUE BRIEF

Evaluating Teachers More Strategically: Using Performance Results to Streamline Evaluation Systems

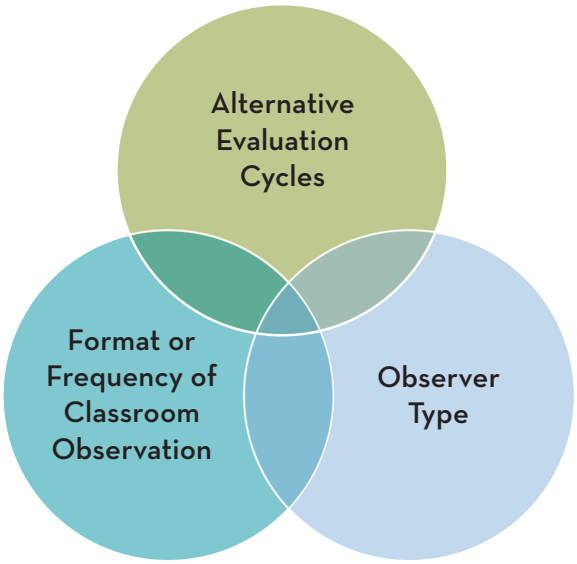
BY TAYLOR WHITE

TEACHER EVALUATION SYSTEMS INTRODUCED by states and school systems in the past several years have focused attention on improving the performance of public school teachers, but they have been cost- and time-intensive, placing a significant burden on states' and districts' resources. In Tennessee, for example, trained evaluators conducted nearly 300,000 classroom observations during the 2011-2012 school year, prompting administrators to complain that "the amount of time spent to implement TEAM [the state's new system] was unmanageable."[1]

Even among school leaders who did not feel overburdened by the number of hours spent observing classrooms, many felt the time had not been used efficiently, as the system treated all teachers with the same intensity, despite the fact that teachers' skills and needs for support varied widely.

To address these concerns, Tennessee and other systems have replaced their one-size-fits-all evaluation approaches with more differentiated models, using past performance data to determine which teachers should be evaluated with more or less intensity in subsequent evaluation cycles and, in some cases, what that attention should include. This brief explores differentiation strategies in nine districts, two charter management organizations, and three states, Tennessee, Delaware, and Ohio.[2] Interviews with system leaders and analyses of teacher evaluation policies reveal that these systems now vary the **format or frequency of formal evaluation cycles,** the **format or frequency of classroom observations,** or the **type of observer** conducting classroom observations, based on what is known about teachers' needs, strengths, and goals.[3]

Many of these school systems have embraced differentiation strategies as a way to conserve teacher evaluation resources or to deploy existing resources more efficiently. In some of the systems, however, differentiation strategies have required *increased* resources, as system leaders have introduced more frequent classroom observation or trained additional observers (e.g., peers, coaches). But even in those instances, officials claim differentiation strategies have helped make evaluation systems more attentive to teachers' individual needs for supervision and support. And that, they say, is likely to lead to more effective teacher evaluation systems.



Common Strategies for Differentiating Teacher Evaluations

Flexibility for Top Performers

Some school systems have sought to shrink principals' workloads by reducing the amount of time they spend evaluating top-performing teachers. In Ohio, for example, districts adopting the state's model evaluation system can opt to evaluate *Accomplished* teachers bi-annually instead of annually, the requirement for teachers rated *Proficient* or below.[4] Delaware's state system allows for a similar approach, although, unlike in Ohio, top-performers are not removed entirely from the evaluation cycle; they receive a "student improvement score" and at least one observation annually, but a full performance evaluation is conducted only once every two years, thereby reducing evaluators' administrative workloads.

Though these approaches can save time, critics worry that such policies may discourage top-performers by limiting their opportunities for feedback and professional growth.[5] To ensure top-performers are still engaged in such opportunities, Burlington, VT; Providence, RI; and Pittsburgh, PA, allow their strongest teachers to complete alternative projects (action research, self-directed study, etc.) in lieu of traditional evaluations. In all three cases, teachers receive annual performance evaluations based on their progress toward specific project goals or outcomes, but the process is still generally less resource-intensive for school administrators, who draw heavily on evidence from teachers' self-reflections and peers' assessments of their progress to determine final evaluative ratings. For these three districts, alternative cycles are also a way to reward top-performers with greater autonomy to design and monitor their own professional growth.

New Formats, Tailored Frequency

Even in school systems that require traditional performance evaluations for all teachers annually, officials now vary *how* and *how often* they observe teachers based on past performance data.

Increasingly, districts are experimenting with different formats for classroom observation, recognizing that traditional, full-length observations may not be the best use of observers' time, or the best way to gather data about teachers' performance. Delaware, for example, now requires districts to use combinations of announced and unannounced classroom observations. Tenured teachers rated *Highly Effective* or *Effective* participate in one announced observation each year, while their lower-performing peers must undergo at least one additional unannounced observation, which provides evaluators a chance to "watch a teacher in action without providing prior notice."[7] The RISE system in Pittsburgh and Tennessee's TEAM system also use combinations of announced and unannounced observations, depending on a teacher's past performance.

Several districts also use walkthroughs—informal check-in visits during which observers spend a brief amount of time in a teacher's classroom focusing on a specific skill or behavior, rather than an entire domain of the observation rubric (as would occur during a full-length observation). This shortened format has become a popular way for observers to gather more frequent and more focused evidence of teacher practice, making it easier for them to provide regular, targeted feedback to teachers on specific areas of need—a task that's difficult to do well when an observer must evaluate several areas of a teacher's practice simultaneously.

The charter school network Achievement First, convinced these shortened observations are the key to driving improvement in teacher practice, has reduced the number of required formal observations to create more time for these shorter, growth-oriented conversations to occur—especially for teachers who need more frequent support. Many other early adopters also note that because walkthroughs are shorter and do not require careful scheduling (most are unannounced), they allow for more frequent interaction between teachers and their observers, a feature which helps build trust and comfort with the process.

Called by different names in different places (e.g., partials, rounds, etc.), versions of these narrowly focused, abbreviated observations have gained popularity in Denver, Providence, Pittsburgh, at the charter network DC Prep, and elsewhere. Though many school systems label these observations *informal,* it's worth pointing out that *informal* does not always indicate that these observations are entirely disconnected from teachers' *formal* performance evaluations. To the contrary, many districts take input from those who conduct informal observations into account in teachers' formal evaluations, though how much weight they carry varies considerably (and is often not formally defined).

> "Traditional, full-length observations may not be the best way to gather data about teachers' performance."

To take advantage of these new observation formats and to use observers' time more strategically, districts and states have also rethought *how often* they observe teachers. In Tennessee, for example, the state board of education recently modified requirements for the

number of observations teachers will undergo under the state's TEAM evaluation system, differentiating the frequency based on a teacher's licensure status *and* his or her last overall evaluation score.

This change, driven by feedback from teachers and administrators after the first year of statewide TEAM implementation, is intended to allow evaluators to "spend more time with the teachers most in need of improvement, while reducing the amount of time spent with teachers whose student outcomes demonstrate strong performance."[6] Teachers in Tennessee now undergo one, two, three, or four full-length observations depending on their prior performance (a composite score of value-added measures and observation scores) and licensure status.

A similar system exists in Hillsborough County, FL, where observers conduct 11 observations of the district's lowest performing teachers, but only three for its highest-performers. And, in Providence, a new Peer Assistance and Review program will provide at least 15 observations and coaching sessions to the district's *Ineffective* or *Developing* teachers, in addition to the observation and support they receive through the district's standard evaluation process.[7]

Other school systems set a minimum number of formal observations, but allow building administrators to conduct additional observations—whether formal or informal—according to teachers' demonstrated needs and the school's available resources. In Jefferson County, CO, for example, teachers must be formally observed three times annually, but schools' Instructional Leadership Teams can opt to conduct between four and ten additional observations based on their assessment of a teacher's performance.[8] Seattle, Denver, Providence, Burlington, Achievement First, DC Prep, and Pittsburgh encourage similar case-by-case differentiation.

New Sets of Eyes

To distribute the responsibility for classroom observations and other evaluation duties, many states and districts now extend observation responsibilities beyond administrators to include high-performing peers, coaches, and other instructional leaders in the evaluation process.[9] Doing so not only reduces principal workloads, but can also provide specialized expertise in particular disciplines (e.g., content, grade-level, or ELL expertise). Using data from prior evaluations, school systems now deploy these observers more strategically, using different combinations of observers for teachers with different needs.

> "Many states and districts now extend observation responsibilities beyond administrators alone, including high-performing peers, coaches, and other instructional leaders."

In Hillsborough County, teachers undergo a specific combination of administrator, peer, and supervisor observations, depending on their evaluation ratings from the previous year. While most teachers in Hillsborough are observed by some combination of administrators and peers, those rated "unsatisfactory" undergo additional supervision by content supervisors.

New Haven mandates a similar combination system, relying on what the district calls "Third Party Validators" (TPVs) to conduct observations of teachers rated at the high or low extremes of the district's scale by their supervisors; TPVs conduct three extra observations for teachers on track to receive *Needs Improvement* ratings and two for those projected to be rated *Exemplary*. By employing these trained external observers, New Haven validates the 'extreme' scores and introduces additional objectivity to the process—all without creating more work for principals and other instructional leaders.

More commonly, policy dictates a minimum number of observations teachers must receive, but provides flexibility for school leaders to determine who conducts those observations based on a teacher's past performance and content-area (or grade-level). In Pittsburgh, for example, administrators and master teachers called Instructional Teacher Leader 2s (ITL2s) share caseloads of "high touch" teachers (pre-tenure and low performers) and "low touch" teachers (average and high-performers with tenure), with the ITL2s taking on more "low touch" cases.[10] Similar discretionary differentiation can be found in Maricopa County, AZ; Jefferson County, Denver, and Providence. Ohio's Teacher Evaluation System (OTES), which also gives school leaders this discretion, goes one step further, giving teachers with above average student growth scores the option to select their own "credentialed evaluator" for formal observations.[11]

In some cases, alternative observers often receive extra compensation (stipends and/or release time) and intensive, on-going training in observation, evaluation, or coaching techniques. Some districts also use external contractors to observe teachers or to score videos of teaching.[12] Though none of the school systems could provide Carnegie cost estimates[13], several reported the addition of alternative observers had been expensive, but, they believe, ultimately effective in providing more supervision and support for teachers.

Realistically, many districts do not have the capacity to build and maintain cadres of alternative evaluators, particularly given how many are struggling to provide sufficient training to principals and other traditional evaluators. And little is known about exactly how to select, train, and deploy these alternative evaluators most effectively or most efficiently. But given that the Gates Foundation's MET project concluded that, "adding a second trained observer increases reliability significantly more than having the same observer score an additional lesson," it seems likely that the use of multiple observers will become more common.[14]

■■■

Though it is too soon to quantify the impact these differentiation strategies have had on teacher evaluation systems (most have been in place for less than three years), evidence from the education systems in this study suggests that they have helped districts deploy their resources more strategically. Even where system leaders have added new components (e.g., training for new types of observers) to their evaluation systems, they report that their investment in more adaptable evaluation systems has allowed them to better match teachers with the supervision and support they need.

	STATES					
	Delaware	Ohio	Tennessee	Burlington, VT	Denver, CO	Hillsborough County, FL
SYSTEM DATA[1]						
Students	131,029	1.87 million	934,000	3,362	78,339	194,525
Schools	216	3,305	1,728	10	158	305
Teachers	8,594	110,000	64,227	343	4,681	13,469
EVALUATION SYSTEM						
System or Framework Title	DPAS II	Ohio Teacher Evaluation System (OTES)	TEAM	Differentiated Teacher Supervision and Evaluation System	LEAP	Empowering Effective Teachers
System-wide Implementation Date	2008-2009	2013	2011-2012	2007-2008	2012-2013	2010-2011
Components of Final Score	Matrix of 5 Components: 4 related to classroom practice, 1 related to student improvement	50% Teacher Performance + 50% Student Performance	35% Student Growth + 15% Academic Achievement + 50% Observation	No Numerical Scores Provided. Qualitative feedback based on observation and conversations between teachers & administrators	30% Observation + 10% Professionalism + 10% Student Perception Data + 50% Student Achievement Data	30% Principal Appraisal + 30% Peer/Mentor Appraisal + 40% Student Achievement Gains
Minimum Number of Formal Observations Required Annually	1 to 3	2	1 to 4	0 to 4	2 to 4	3 to 11
Observation Formats Used	Formal Announced & Unannouced	30 min. Formal & Informal Walkthroughs	Formal Announced & Unannounced, Walkthroughs	Formal and Informal	45-60 min. Formal, 20 min. Partials, 10 min. Walkthroughs, Peer	Admin. Formal & Informal, Peer Formal & Informal, Supervisor Formal
Effectiveness Levels	4	4	5	3	4	4
WHO CONDUCTS OBSERVATIONS?						
Administrators	Yes	Yes	Yes	Yes	Yes	Yes
Alternative Observers (Peers, Mentors, Coaches, Third Party, etc.)	Yes	Yes	District discretion	Yes	Yes	Yes
BASED ON PRIOR EVALUATION RESULTS, DOES THE SYSTEM DIFFERENTIATE....						
Freqency or type of evaluation cycle?	Yes	Yes	No	Yes	No	No
Frequency or type of mandatory classroom observations?	Yes	No[8]	Yes	Yes	No	Yes
Observer Type?	No	District discretion	District discretion	Yes	No	Yes

[1] System data from NCES (2010-2011) or from CMO leadership.

[2] Data in this table describes a pilot program in Jefferson County, CO. The pilot, funded by TIF, impacts the number of students, schools, and teachers in parenthesis.

[3] The REIL project directed by the Maricopa County Education Service Agency includes six districts (Alhambra, Gila Bend, Isaac, Nadaburg, Phoenix Union, and Tolleson) in greater Phoenix, AZ.

[4] Data for Providence, Rhode Island represents policies set by the Rhode Island Innovation Consortium, of which Providence is a member. Some of these policies have been provisionally approved but have not yet been enacted. For more information, see the profile for the district.

[5] In Pittsburgh, high-performing teachers doing Supported Growth Projects in lieu of the formal RISE process are not observed formally while in a project year.

DISTRICTS						CHARTERS	
Jefferson County, CO[2]	Maricopa County, AZ[3]	New Haven, CT	Pittsburgh, PA	Providence, RI[4]	Seattle, WA	Achievement First (NY, CT, RI)	DC Prep (DC)
85, 979 (8,550)	52,294	20,003	27,982	23,573	47,735	7,000	1,200
166 (20)	52	48	69	51	103	22	4
4,869 (155)	3,380	1,571	2,076	1,590	5,584	600	113
TLL	REIL	TEVAL	Empowering Effective Teachers / Research-based Inclusive System of Evaluation (RISE)	Education Evaluation & Support System (RIIC)	PG&E	Teacher Career Pathway	Continuous Improvement Cycle
2012-2013	2010-2011	2009-2010	2010-2011	2010-2011	2012-2013	2011-2012	2010-2011
Matrix: 50% Professional Practices + 50% Student Academic Growth (10% District Goal + 15% School Goal + 10% School Performance Framework Rating + 15% Individual Education Goals)	50% Observation + 40% Individual Growth + 5% Team Growth + 5% School Growth	Matrix: Instructional Practice and Professional Values Score & Student Learning Growth Score	50% Observation + 30% Student Growth + 15% Student Survey Data + 5% Building-level test scores	Matrix: Professional Practice + Professional Growth & Responsibilities + Student Learning Rating (SLO + Growth Score (where applicable))	Observations + Student Learning Objectives	Student Outcomes (Achievement Scores + Student & Parent Surveys) + Teacher Inputs (Observations + Peer & Principal Survey)	Classroom Observation + Lesson plan feedback & submission rates + Student Data (including, but not limited to achievement data, character development data, attendance data, etc.)
3	5	1	0 to 5[5]	0 to 1[6]	2 to 3	2	1
Formal & Informal (4-10 mandatory)	Principal & Peer Formal, Informal/ Walkthrough (varies by district)	Formal and Mandatory Instructional Rounds Observations	Formal & Informal, Announced & Unannounced	30-60 min. Formal (Announced) & 10-20 min. Informal (Unannounced)	30 minute Formal & Informal	Formal and Informal	Formal and Informal
4	5	5	4	4	4 (for Obs)/5 (for SLO)	5	4
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes	No[7]	Yes	Yes
No	No	No	Yes	Yes	Yes	No	No
No[8]	No	Yes	Yes	Yes	No[8]	No[8]	No[8]
Yes	No	Yes	Yes	Yes	No	Yes	Yes

[6] Tenured teachers in Providence who receive a Highly Effective Rating in their PPG&R domain can participate in the district's Differentiated Model of evaluation which requires two informal observations, but no formal observation.

[7] Seattle's STAR program uses veteran teachers as mentors for novice teachers, but their observations only supplement those conducted by an administrators; they are not formal observations.

[8] These districts set minimum requirements for annual formal evaluations but allow school-level leadership to conduct additional observations at their discretion. The frequency and format of these additional observations varies across districts.

ENDNOTES

[1] Tennessee Board of Education. (2012). Teacher Evaluation in Tennessee: A Report on Year 1 Implementation: 20.

[2] This brief considers the model evaluation systems created by the state departments of education in Tennessee (TEAM) and Ohio, but recognizes that other district-specific systems are in place in those states. Delaware's DPAS II's system is mandatory for all districts.

[3] Many districts also provide additional support for novice teachers, regardless of their evaluation results. Though these programs are promising strategies for improving practice and retaining teachers, this brief does not include such programs because they tend to differentiate based on experience rather than performance. Additionally, most are supportive, rather than evaluative in nature.

[4] Accomplished and Proficient are rating categories used by the Ohio Department of Education. Here and throughout the brief, systems' terminology will be used when describing rating categories.

[5] Ohio's state model allows for alternative evaluation formats for top-performers. It is not clear how many districts have chosen or will choose to introduce such formats.

[6] Delaware Department of Education. (2013). DPAS II: Revised Guide for Teachers: 50.

[7] Providence's PAR program provides similarly intensive support to the district's novice teachers and to veteran teachers new to the district. The program's purpose is to provide support to improve teachers' practice, rather than to provide more opportunities for evaluative observation. Seattle's STAR program provides similarly intense, growth-oriented mentoring for novice teachers as well.

[8] This policy applies only to the 20 schools participating in the Jefferson County Public Schools' Teacher Incentive Fund Pilot project.

[9] For more on the impact of multiple observers, see Ho, A.D. and Kane, T.J. (2013). The Reliability of Classroom Observations by School Personnel. Bill & Melinda Gates Foundation.

[10] ITL2s cannot issue final formal evaluations of teachers, however, even if their observations are a part of that score. In Pittsburgh and many other school systems throughout the country, only administrators may issue summative ratings.

[11] Teachers with average student growth scores can "have input" in selecting their evaluator; those with below average scores have no say and are assigned an evaluator.

[12] Teachscape provided this service to the Measures of Effective Teaching project and serves as a clear example of how this might work. See McClellan et. al. (2012).

[13] The Carnegie Foundation has developed an online cost calculator to help district employees and members of the K-12 community understand the different components of designing a district's teacher evaluation system. For more information, please visit: http://commons.carnegiefoundation.org/what-we-are-learning/2013/carnegie-cost-calculator-a-tool-for-exploring-the-cost-of-educator-evaluation-systems/

[14] The Bill & Melinda Gates Foundation. (2013) "Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study.": 5.

INTERVIEWS AND PERSONAL CORRESPONDENCE

- David R. Bell, Director, HRIS Compensation & Records, Jefferson County Public Schools
- Beth Berwick, Chief of Staff, DC Prep
- Anna Brown, Director for Assessment and Performance Management, Empowering Effective Teachers Grant, Hillsborough County Public Schools
- Maribeth Calabro, VP At-Large, Providence Teachers Union
- Clover Codd, Director of the Teacher Incentive Fund, Seattle Public Schools
- Sarah Coon, Former Senior Director of Talent Development, Achievement First
- Justin Darnell, Senior Manager of PD Alignment, Denver Public Schools
- Samuel Franklin, Executive Director, Office of Teacher Effectiveness, Pittsburgh Public Schools
- Errika Jones, Coordinator, Office of Teacher Effectiveness, Pittsburgh Public Schools
- Luke Kohlmoos, Director of TEAM, Tennessee Department of Education
- Sarah Mahan, Director of Human Resources, Burlington School District (VT)
- Kristy Parsons, Assistant Director, Teacher Incentive Fund, Jefferson County Public Schools and Jefferson County Education Association Representative, Board of Directors, Colorado Education Association
- Matthew Pepper, Research and Data Quality Coordinator, National Center on Performance Initiatives at Vanderbilt University
- Lori Renfro, Assistant Superintendent, Performance Management Systems, Maricopa County Education Service Agency
- Alicia Robinson, former Chief Talent Officer, DC Prep
- Christopher Ruszkowski, Chief Officer, Teacher Leadership and Effectiveness Unit, Delaware Department of Education
- Michele Sherban-Kline, Director, Teacher Evaluation and Development, New Haven Public Schools
- Julia Simmerer, Director, Ohio Department of Education Office of Educator Effectiveness
- Kendra Wilhelm, LEAP Program Manager, Denver Public Schools
- Maureen Yoder, Associate Director, Ohio Department of Education Office of Educator Effectiveness

SOURCES

Archibald, S., Coggshall, J.G., Croft, A. & Goe, L. (2011). *High-Quality Professional Development for All Teachers: Effectively Allocating Resources.* National Comprehensive Center for Teacher Quality.

Burlington School District. (2007). *Differentiated Teacher Supervision and Evaluation System.*

Delaware Department of Education. (2013). *DPAS II: Revised Guide for Teachers.* Retrieved from: http://www.doe.k12.de.us/csa/dpasii/files/teachforms/DPASTeachFullGuide.pdf

Doyle, D., & Han, J.G. (2012). *Measuring teacher effectiveness: A look "under the hood" of teacher evaluation in 10 sites.* New York: 50CAN; New Haven, CT: ConnCAN; and Chapel Hill, NC: Public Impact.

Goe, L., Biggers, K., & Croft, A. (2012). *Linking Teacher Evaluation to Professional Development: Focusing on Improving Teaching and Learning.* National Comprehensive Center for Teacher Quality.

Hillsborough County Public Schools. (2011). *Teacher Evaluation Handbook: Empowering Effective Teachers.* Retrieved from: http://www.fldoe.org/profdev/pdf/pa/Hillsborough.pdf

Ho, A.D. and Kane, T.J. (2013). The Reliability of Classroom Observations by School Personnel. Bill & Melinda Gates Foundation.

McClellan, C., Atkinson, M., Danielson, C. (2012). Teacher Evaluator Training & Certification: Lessons Learned from the Measures of Effective Teaching Project. Practitioner Series for Teacher Evaluation. Teachscape. San Francisco, CA.

Rhode Island Federation of Teachers and Health Professionals. (2012). Rhode Island Innovation Consortium Education Evaluation & Support System: RIDE Approved for Full-Implementation.

Tennessee Board of Education. (2012). *Teacher Evaluation in Tennessee: A Report on Year 1 Implementation.* Retrieved from: https://www.tn.gov/education/doc/yr_1_tchr_eval_rpt.pdf

The Bill and Melinda Gates Foundation. (2013). Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study.

Carnegie Foundation
for the Advancement of Teaching

Carnegie Foundation for the Advancement of Teaching
51 Vista Lane
Stanford, California 94305
650-566-5100

Carnegie Foundation for the Advancement of Teaching is committed to developing networks of ideas, individuals, and institutions to advance teaching and learning. We join together scholars, practitioners, and designers in new ways to solve problems of educational practice. Toward this end, we work to integrate the discipline of improvement science into education with the goal of building the field's capacity to improve.

www.carnegiefoundation.org

This publication was written with funding from The Joyce Foundation. The statements made and views expressed are solely the responsibility of the author. Carnegie is also grateful to the Institute of Education Sciences for supporting the Foundation's work on teacher evaluation and the improvement of teaching.